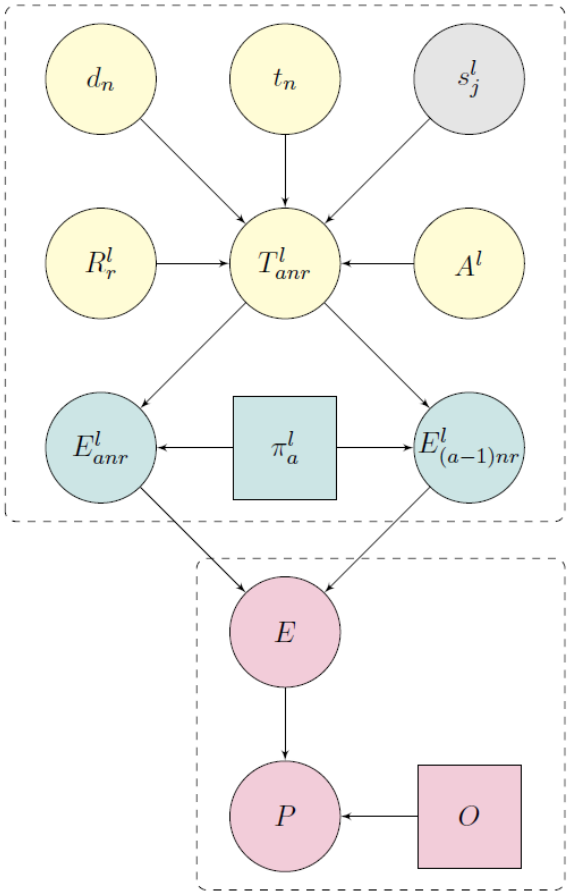
Likelihood Ratio Modeling

Continuous Model Network

The continuous model we are going to discuss consists of several elements:



Adapted from: The interpretation of single source and mixed DNA profiles (Taylor et al., 2013).

LR Modeling

The LR can now be assessed by writing the ratio in the form:

$$\mathsf{LR} = \frac{\mathsf{Pr}(G_C|G_S, H_p, I)}{\mathsf{Pr}(G_C|G_S, H_d, I)}$$

$$= \frac{\sum_j \mathsf{Pr}(G_C|S_j)\,\mathsf{Pr}(S_j|H_p)}{\sum_{j'} \mathsf{Pr}(G_C|S_{j'})\,\mathsf{Pr}(S_{j'}|H_d)}$$

$$= \frac{\sum_j w_j\,\mathsf{Pr}(S_j|H_p)}{\sum_{j'} w_{j'}\,\mathsf{Pr}(S_{j'}|H_d)}.$$

The two propositions each define sets of genotypes S, and the weights w describe how well these sets fit our observed data G_C. Under H_p all the genotype sets S_j usually include G_S.

Modeling Strategies

Now that a model has been developed, we require information about the input parameters.

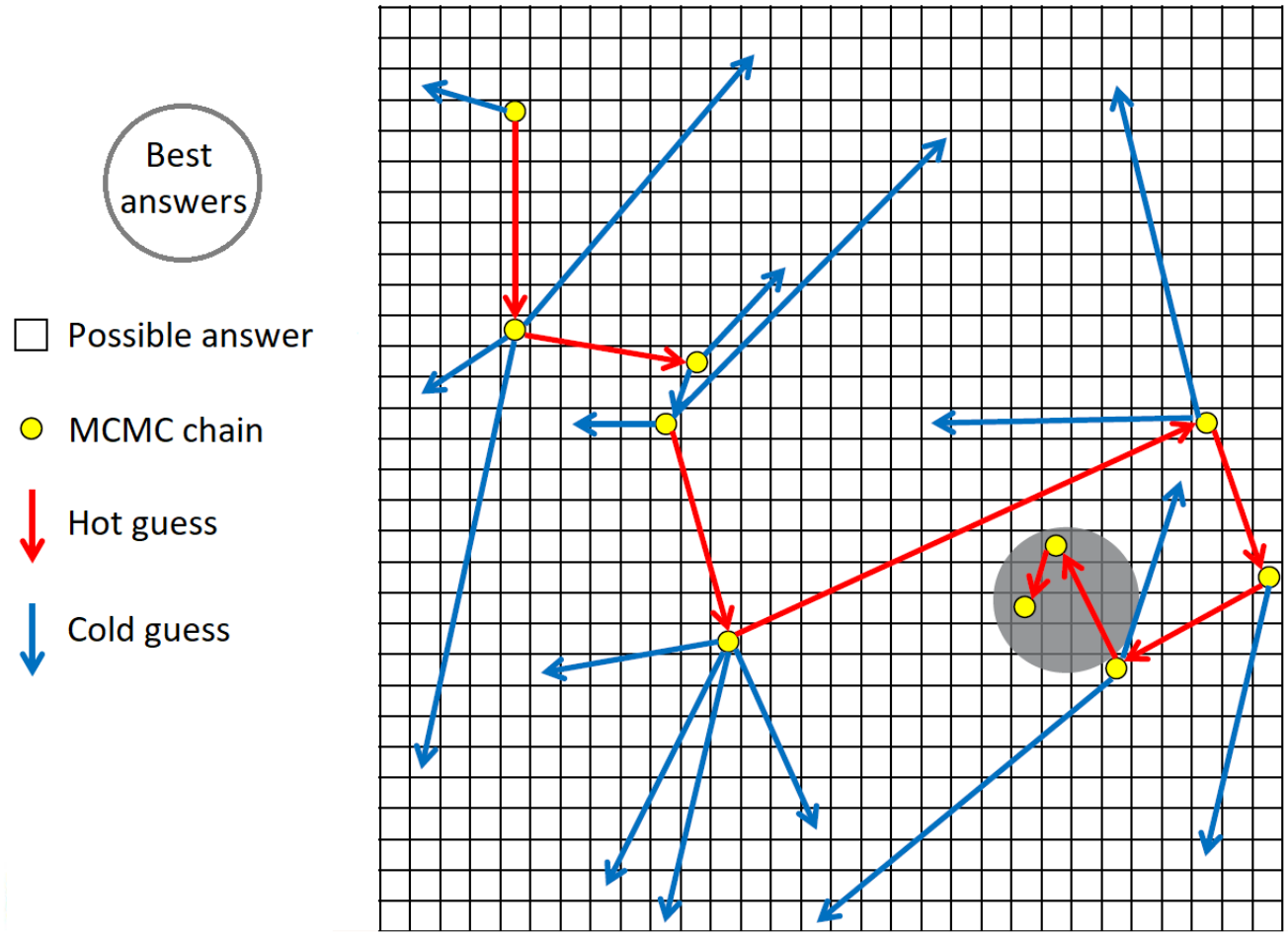
- **Maximization**: Parameters can be chosen that maximize the likelihood of the observations under each hypothesis.

- **Integration**: Rather than knowing the true values of the parameters, we need to know the effect they have on the probability of the observed data.

- **Markov chain Monte Carlo**: Instead of testing every possible combination of parameters, only a small distribution of parameter values and genotype sets will accurately describe the data.

Markov chain Monte Carlo (MCMC)

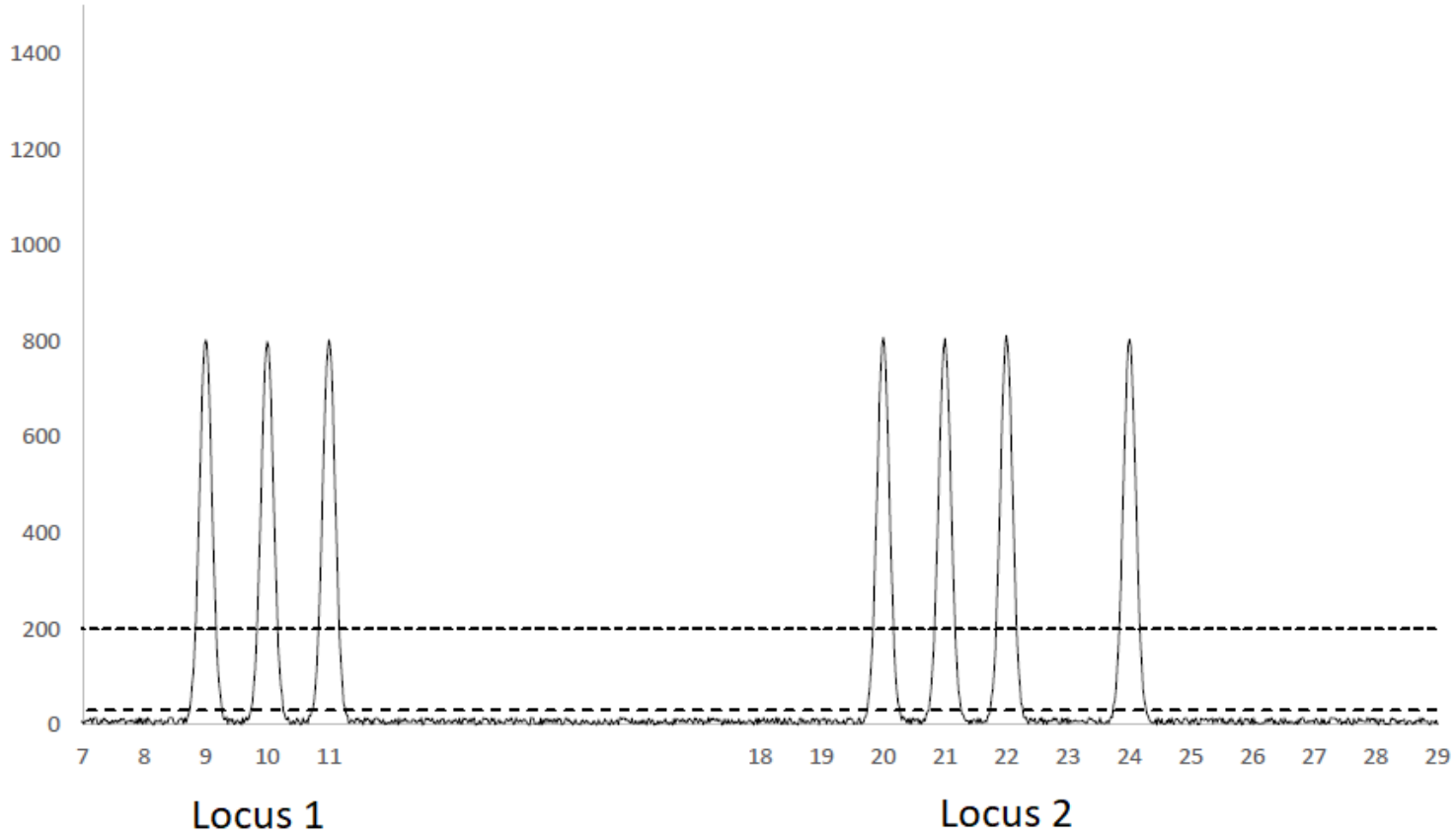
MCMC will start by choosing parameter values at random, eventually leading to more sensible options, until it has reached an equilibrium state.



Expected Peak Heights

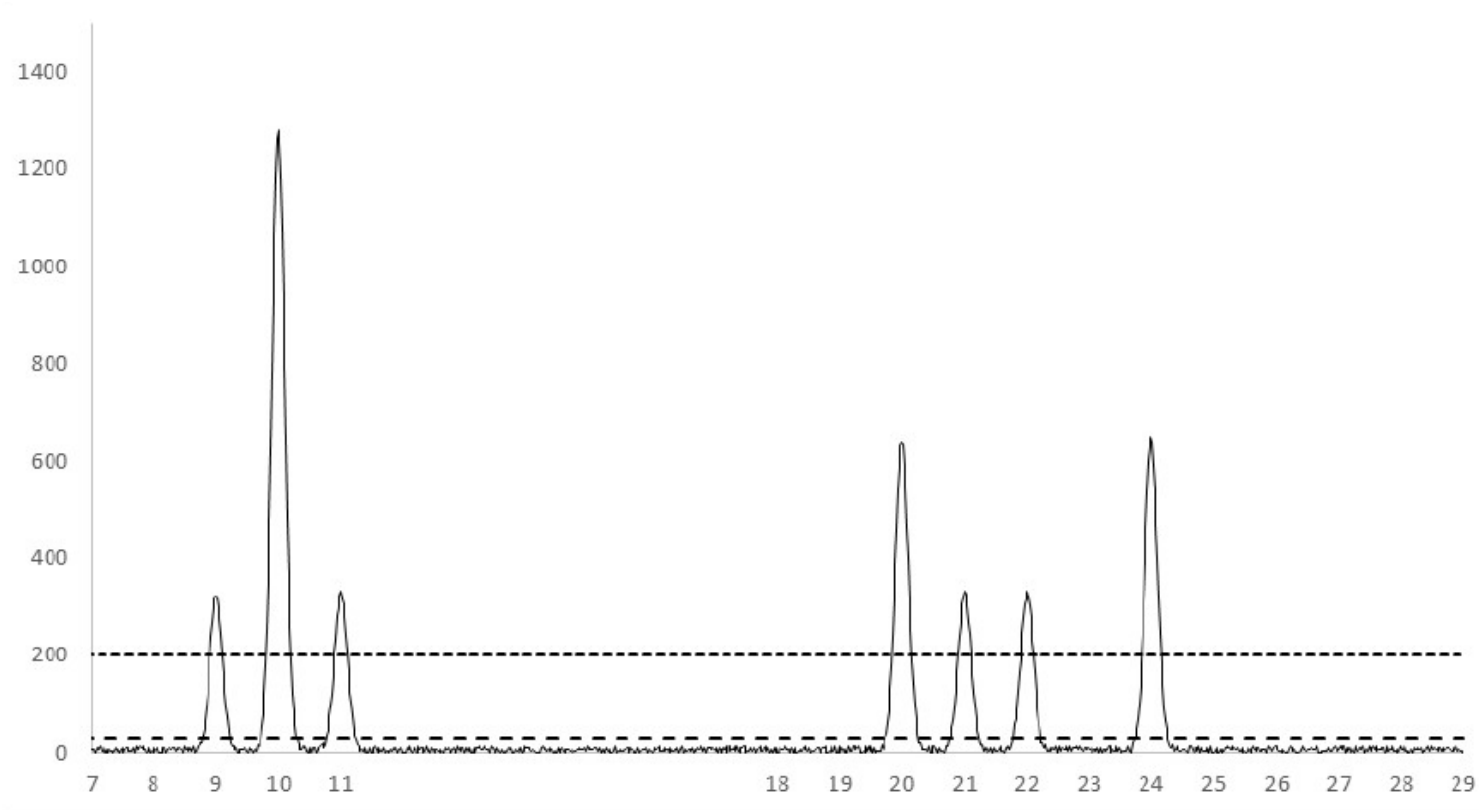
Based on a set of input parameters, an expected profile can be generated.

Step 1: Genotypes are chosen.



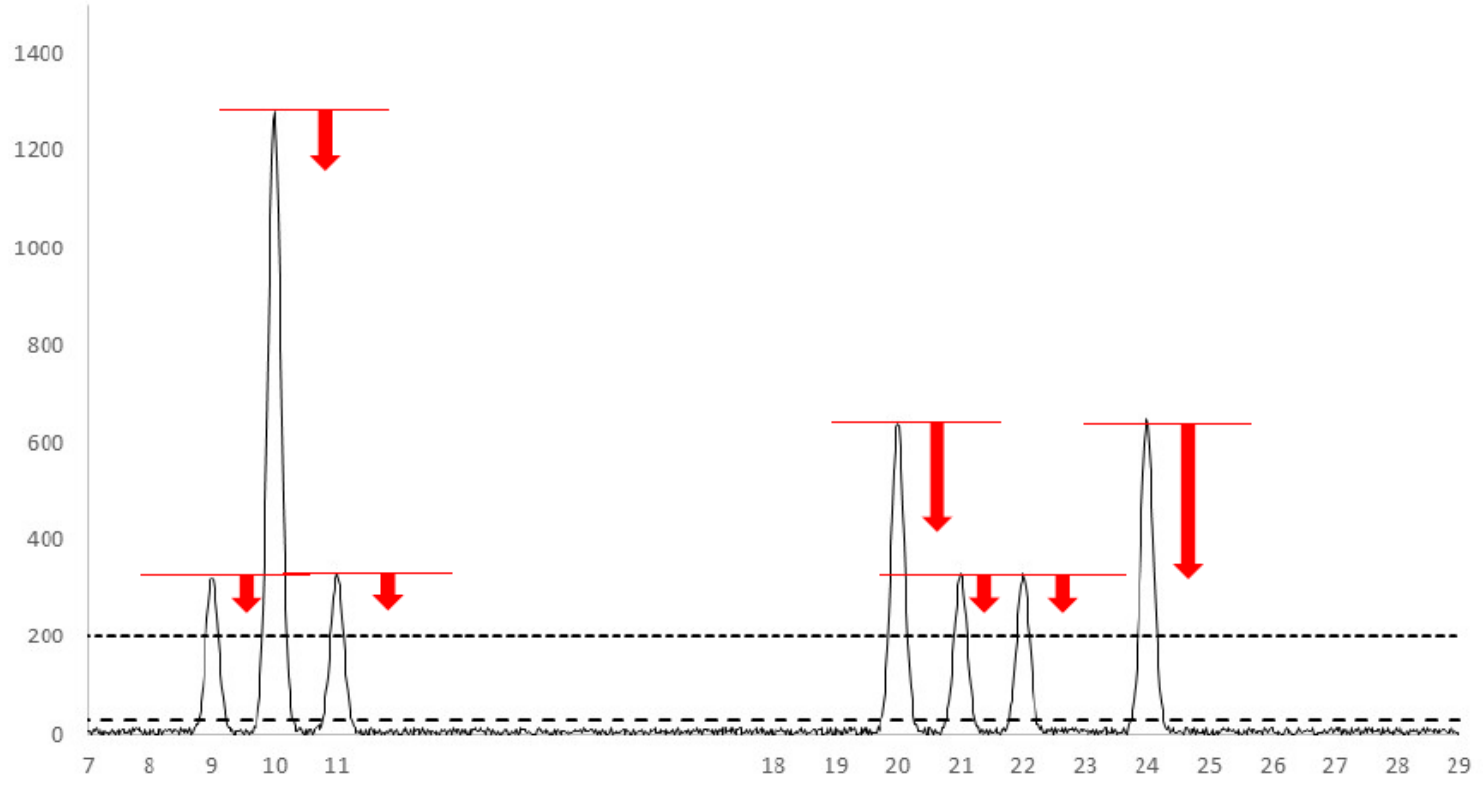
Expected Peak Heights

Step 2: Template amounts per contributor are incorporated.



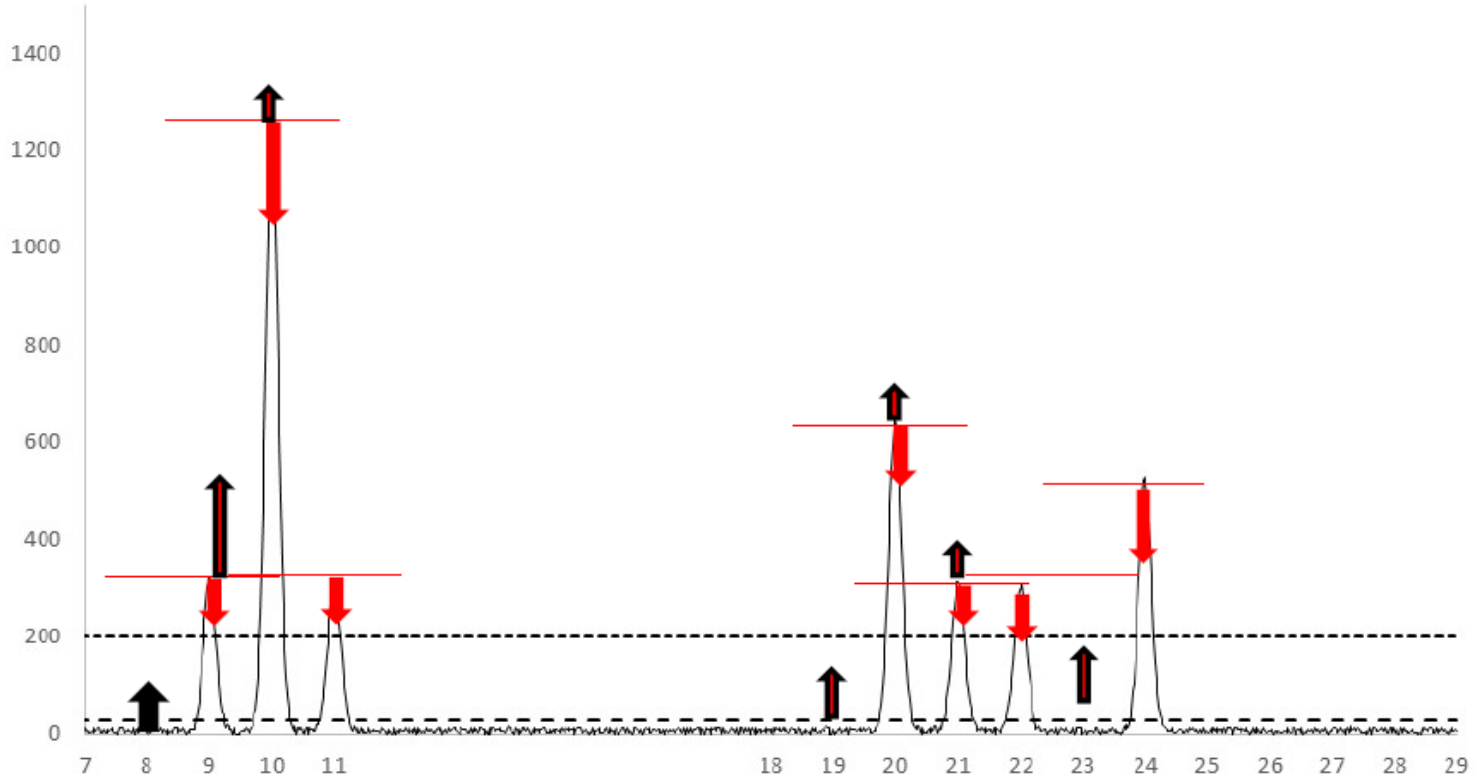
Expected Peak Heights

Step 3: Degradation is taken into account.



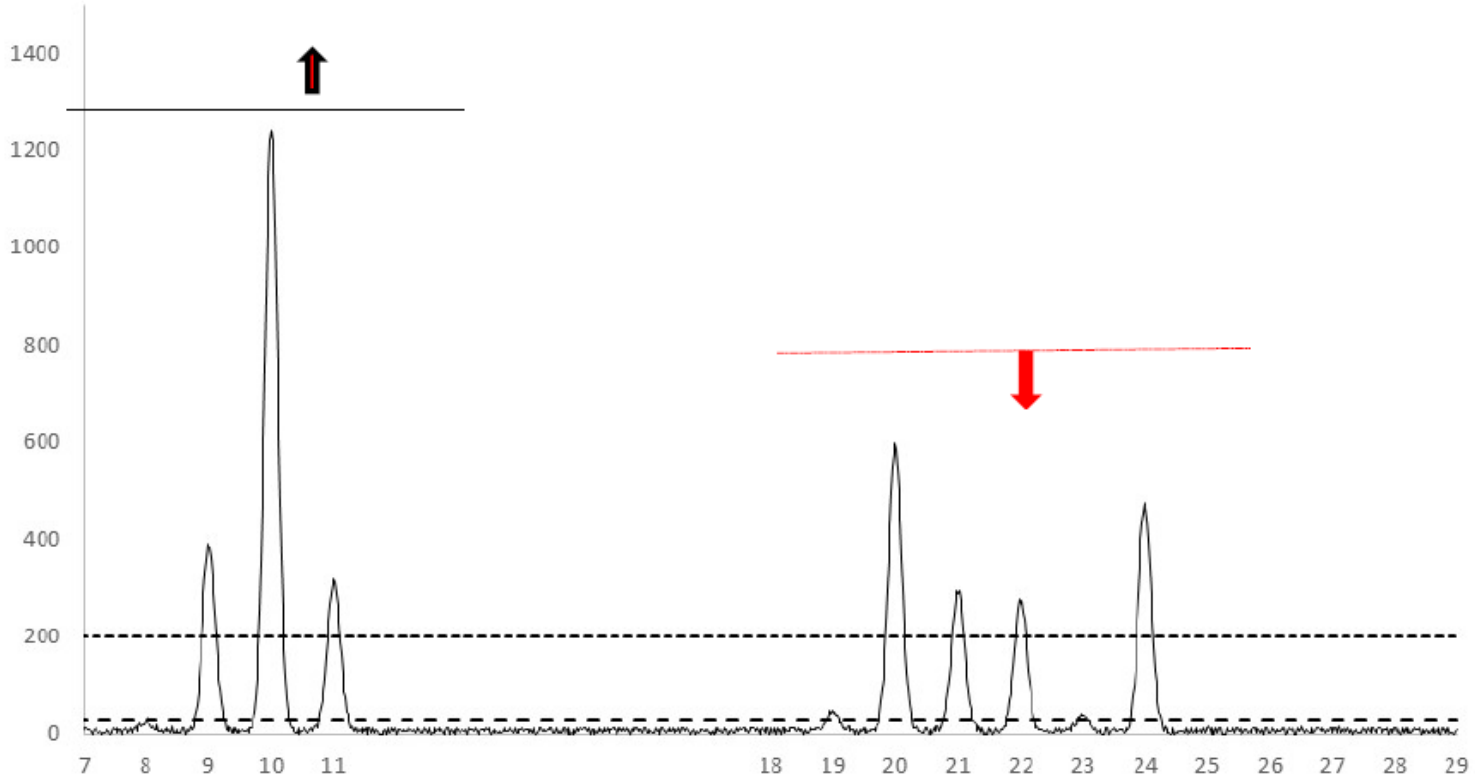
Expected Peak Heights

Step 4: Stutter is taken into account.



Expected Peak Heights

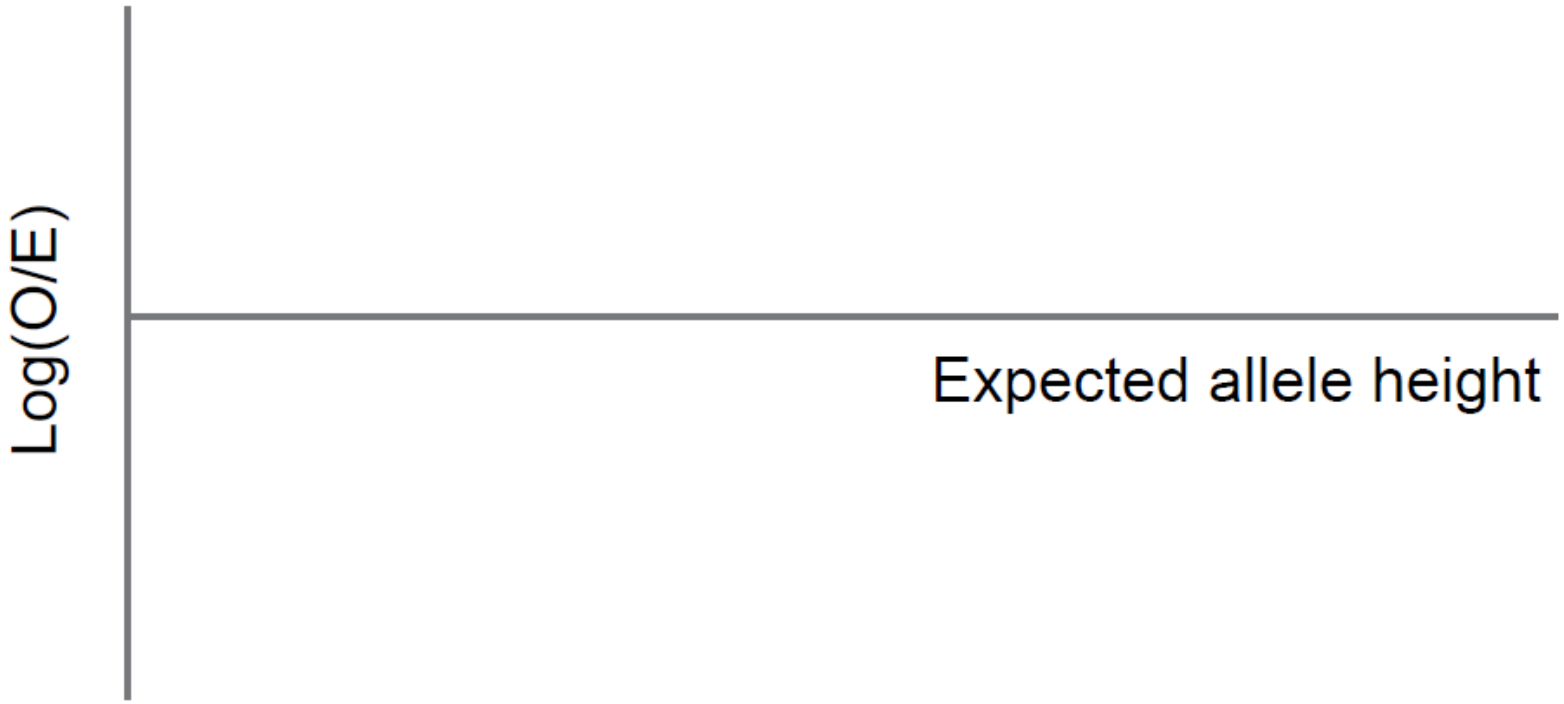
Step 5: Locus specific amplification efficiencies are introduced.



The Perfect Model

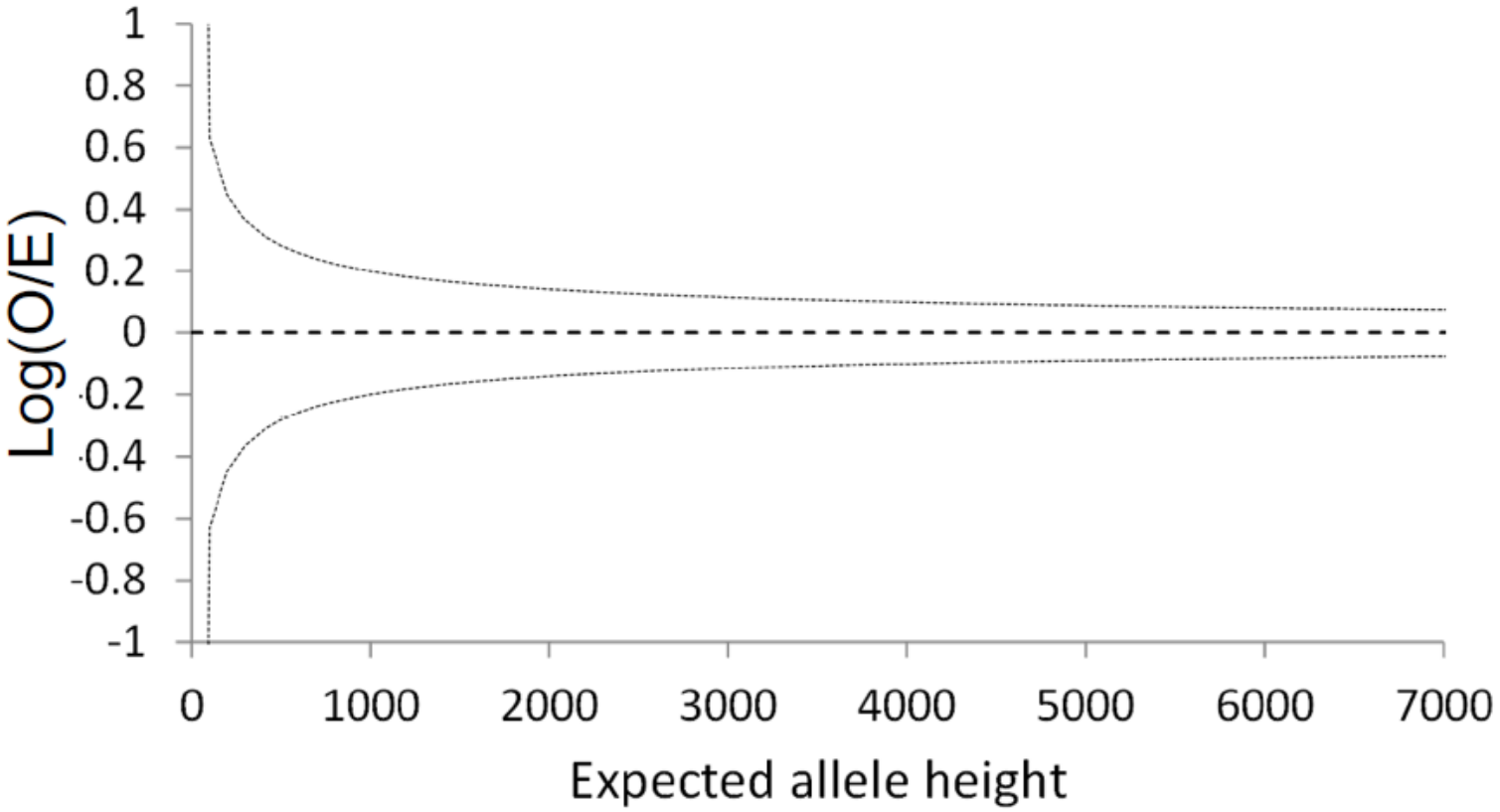
We can now compare our expected profile with the observed STR profile.

What would a perfect model look like?



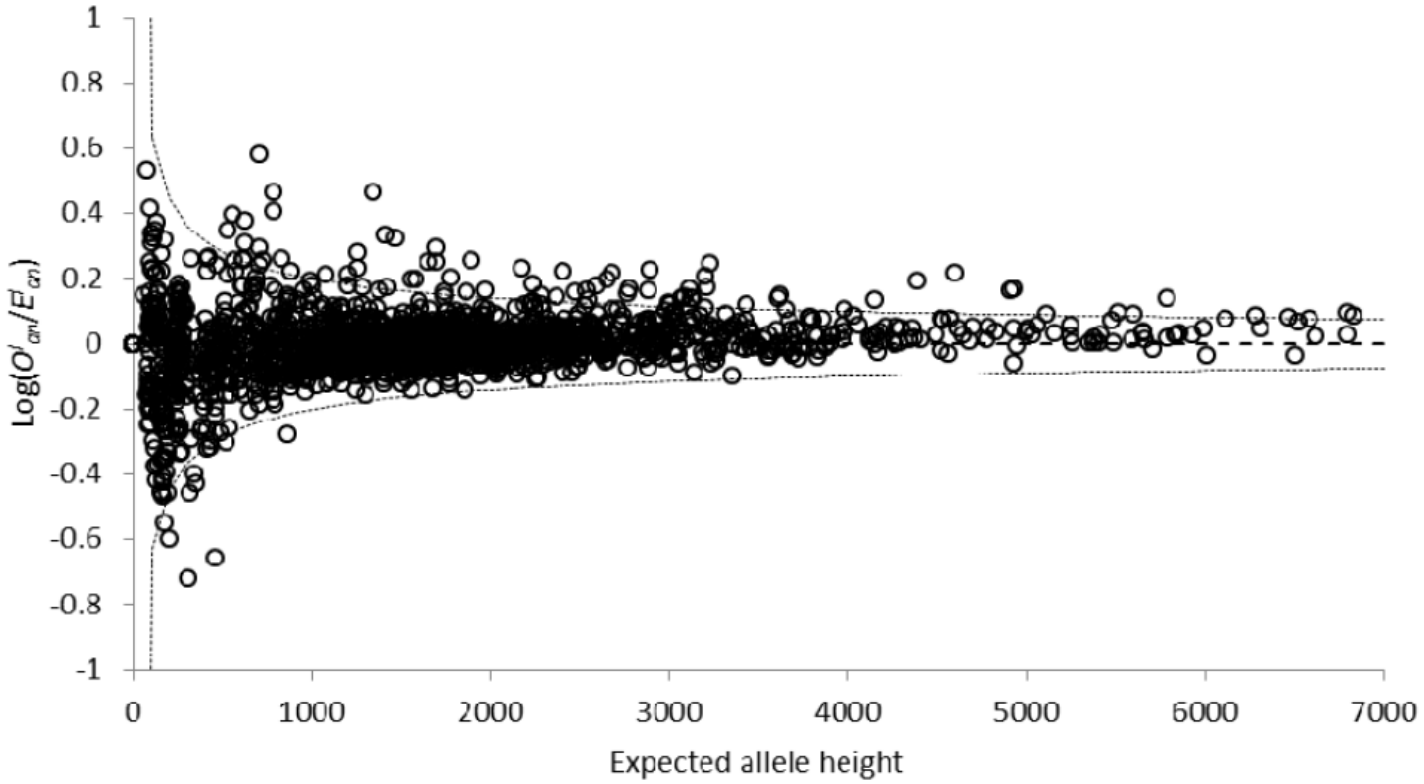
The Perfect Model

Observations show that the relative variance of small peaks is large and the relative variance of large peaks is small. This suggests that the variance is inversely proportional to the expected peak height.



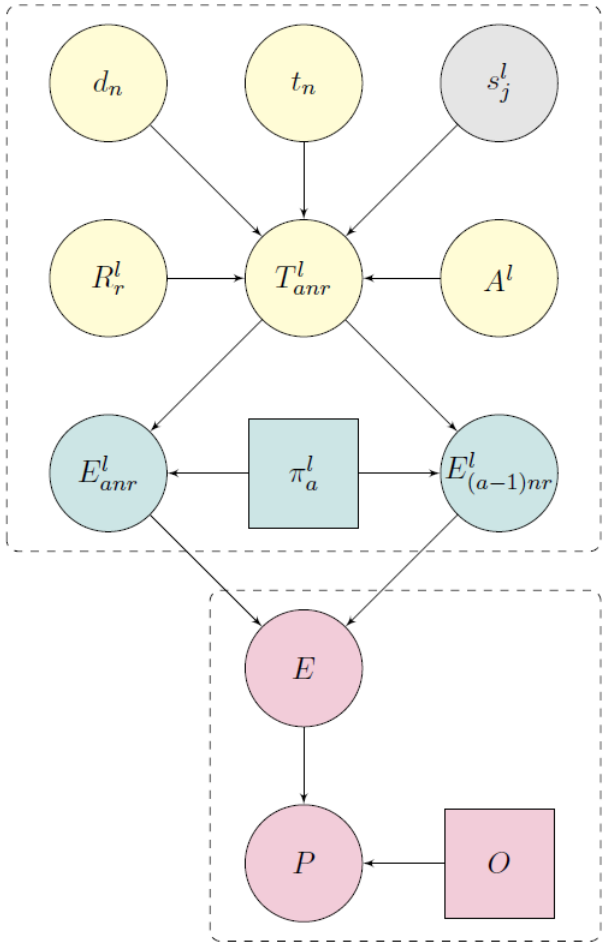
Generating Weights

The probability of obtaining the observed profile can now be calculated by considering the ratio of the observed and expected peak heights, assuming the log of this ratio has mean 0 and variance proportional to $1/E$.



Continuous Model Network

Combining all elements leads to an overall continuous model network:



Source: The interpretation of single source and mixed DNA profiles (Taylor et al., 2013).

Available Software

Not all models as published in literature have been translated into software. A non-exhaustive list:

Software	Class	Availability	Optimization
LRmix Studio	semi-continuous	open-source	ML
Lab Retriever	semi-continuous	open-source	ML
MixKin	semi-continuous	in-house	Integration
DNA LiRA	(semi-)continuous	open-source	Bayes
likeLTD	(semi-)continuous	open-source	ML
STRmix	continuous	commercial	Bayes
TrueAllele	continuous	commercial	Bayes
DNA·VIEW	continuous	commercial	ML
DNAmixtures	continuous	open-source*	ML
EuroForMix	continuous	open-source	ML or Bayes
DNAStatistX	continuous	in development	ML

See also: Probabilistic Genotyping Software: An Overview (Coble & Bright, 2019).